

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

JPA 11-205451

(11) Publication number: 11205451 A

(43) Date of publication of application: 30.07.99

(51) Int. Cl.  
H04M 3/42  
G10L 3/00  
G10L 3/00  
G10L 5/06

(21) Application number: 10007901

(71) Applicant: CANON INC

(22) Date of filing: 19.01.98

(72) Inventor: YAMADA MASAOKI

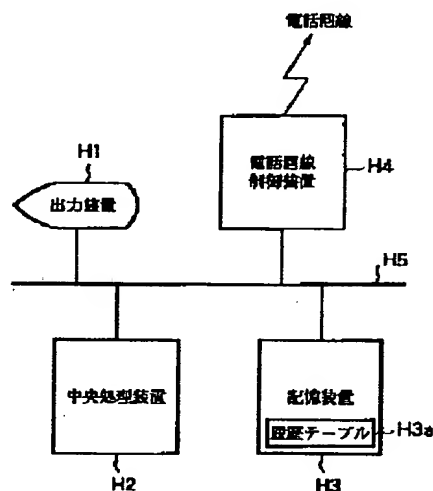
(54) SPEECH RECOGNITION DEVICE, METHOD THEREFOR AND COMPUTER READABLE MEMORY

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a speech recognition device capable of improving recognition accuracy, its method and a computer readable memory.

SOLUTION: A history table H3a for managing caller information relating to a caller for each telephone number is stored and the telephone number of the caller is obtained from a telephone line controller H4. The caller information corresponding to the obtained telephone number is retrieved from the history table H3a. By using the retrieved caller information and voice data originated from the caller, in a central processing unit H2, a recognition environment is adapted. Then, based on the adapted recognition environment, the voice data are recognized.

COPYRIGHT: (C)1999,JPO



**THIS PAGE BLANK (USPTO)**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-205451

(43) 公開日 平成11年(1999) 7月30日

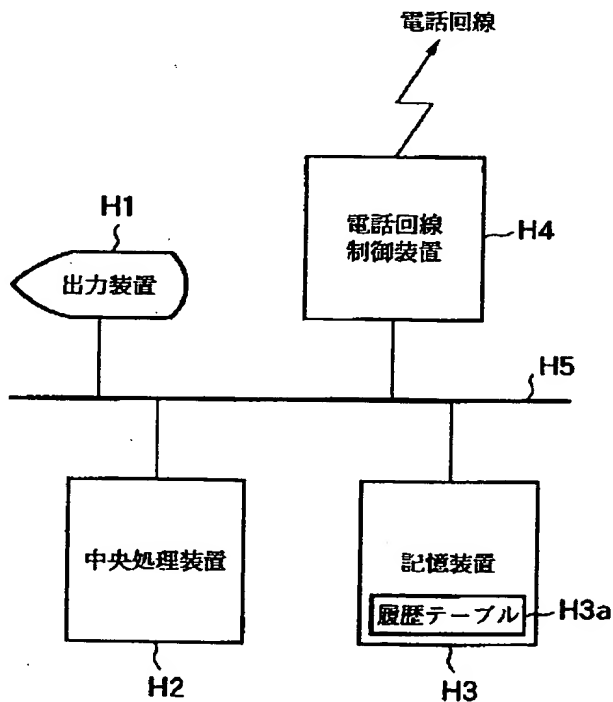
(51) Int. Cl. <sup>6</sup>	識別記号	F I			
H04M 3/42		H04M 3/42		P	
				D	
G10L 3/00	531	G10L 3/00	531	L	
	551		551	A	
5/06		5/06		Z	
		審査請求	未請求	請求項の数13	OL (全6頁)
(21) 出願番号	特願平10-7901	(71) 出願人	000001007		
			キヤノン株式会社		
			東京都大田区下丸子3丁目30番2号		
(22) 出願日	平成10年(1998) 1月19日	(72) 発明者	山田 雅章		
			東京都大田区下丸子3丁目30番2号		
			キヤノン株式会社内		
		(74) 代理人	弁理士 大塚 康德 (外2名)		

(54) 【発明の名称】 音声認識装置及びその方法、コンピュータ可読メモリ

(57) 【要約】

【課題】 認識精度を向上することができる音声認識装置及びその方法、コンピュータ可読メモリを提供する。

【解決手段】 電話番号毎に発信者に関する発信者情報を管理する履歴テーブルH3aを記憶し、発信者の電話番号を電話回線制御装置H4より取得する。取得した電話番号に対応する発信者情報を履歴テーブルH3aより検索する。検索された発信者情報及び前記発信者から発信された音声データを用いて、中央処理装置H2において、認識環境を適応する。そして、適応された認識環境に基づいて、音声データを認識する。



## 【特許請求の範囲】

【請求項1】 電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識装置であって、

電話番号毎に発信者に関する発信者情報を管理する管理テーブルを記憶する記憶手段と、

前記発信者の電話番号を取得する取得手段と、

前記取得手段で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索手段と、

前記検索手段で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適応する適応手段と、

前記適応手段で適応された認識環境に基づいて、前記音声データを認識する認識手段とを備えることを特徴とする音声認識装置。

【請求項2】 前記記憶手段は、前記取得手段で電話番号が取得される毎に、前記管理テーブルの該電話番号に対応する発信者情報を更新することを特徴とする請求項1に記載の音声認識装置。

【請求項3】 前記発信者情報は、少なくとも音声データ、該音声データに信号処理を施したパラメータ列、統計的情報のいずれかあるいは任意の組み合わせを含むことを特徴とする請求項1に記載の音声認識装置。

【請求項4】 前記記憶手段は、前記発信者情報として、更に、該発信者情報の作成日を前記管理テーブルに記憶することを特徴とする請求項3に記載の音声認識装置。

【請求項5】 前記記憶手段は、前記発信者情報として、更に、発信者が使用する電話器の種類を前記管理テーブルに記憶することを特徴とする請求項3に記載の音声認識装置。

【請求項6】 前記パラメータ列は、少なくともパワースペクトル、ケプストラム、LPCケプストラム、MFCCのいずれかを含むことを特徴とする請求項3に記載の音声認識装置。

【請求項7】 電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識方法であって、

電話番号毎に発信者に関する発信者情報を管理テーブルに記憶する記憶工程と、

前記発信者の電話番号を取得する取得工程と、

前記取得工程で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索工程と、

前記検索工程で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適応する適応工程と、

前記適応工程で適応された認識環境に基づいて、前記音声データを認識する認識工程とを備えることを特徴とする音声認識方法。

【請求項8】 前記記憶工程は、前記取得工程で電話番

号が取得される毎に、前記管理テーブルの該電話番号に対応する発信者情報を更新することを特徴とする請求項7に記載の音声認識方法。

【請求項9】 前記発信者情報は、少なくとも音声データ、該音声データに信号処理を施したパラメータ列、統計的情報のいずれかあるいは任意の組み合わせを含むことを特徴とする請求項7に記載の音声認識方法。

【請求項10】 前記記憶工程は、前記発信者情報として、更に、該発信者情報の作成日を前記管理テーブルに記憶することを特徴とする請求項9に記載の音声認識方法。

【請求項11】 前記記憶工程は、前記発信者情報として、更に、発信者が使用する電話器の種類を前記管理テーブルに記憶することを特徴とする請求項9に記載の音声認識方法。

【請求項12】 前記パラメータ列は、少なくともパワースペクトル、ケプストラム、LPCケプストラム、MFCCのいずれかを含むことを特徴とする請求項9に記載の音声認識方法。

【請求項13】 電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識のプログラムコードが格納されたコンピュータ可読メモリであって、

電話番号毎に発信者に関する発信者情報を管理テーブルに記憶する記憶工程のプログラムコードと、

前記発信者の電話番号を取得する取得工程のプログラムコードと、

前記取得工程で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索工程のプログラムコードと、

前記検索工程で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適応する適応工程のプログラムコードと、

前記適応工程で適応された認識環境に基づいて、前記音声データを認識する認識工程のプログラムコードとを備えることを特徴とするコンピュータ可読メモリ。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識装置及びその方法、コンピュータ可読メモリに関するものである。

【0002】

【従来の技術】従来より、音声認識の手法として、背景雑音・入力回線・マイクロホン・話者といった認識環境に適応する技術が存在する。

【0003】例えば、背景雑音に適応する技術としては、スペクトルサブトラクション(SS)法あるいは雑音モデル重畳(PMC)法等がある。SS法では、入力された音声データのうち無音声部分のパワースペクトル

を求め、その平均値を入力された音声データから差し引いたのち認識を行う。PMC法では、無音声部分からノイズを表現するモデルを作成し、認識に用いるモデルと融合させて用いる。

【0004】また、入力回線・マイクロホンに適應する技術として、平均ケプストラムサブトラクション(CMS)法がある。これは、入力された音声データのケプストラムの平均を求め、入力された音声データからそのケプストラムを差し引いたのち認識を行うものである。これと等価な方法として平均ケプストラムを認識に用いるモデルに足し込む方法もある。

【0005】また、話者に適應する技術として、認識に用いるモデルに対する尤度が最大になるようにモデルパラメータを更新する方法や、モデル作成時に話者をクラスタリングし、認識時に最適な話者クラスタを選択する方法がある。

【0006】また、上記技術の組合せとして、様々な認識環境に適應する技術が提案・実施されている。

【0007】

【発明が解決しようとする課題】しかしながら、上記従来例では、次のような問題点があった。

【0008】認識環境の適應に用いるための情報を示すデータ量が少ない場合、通常の適應に用いるためのデータ量が不足して発生する過適應によって動作が不安定になり、かえって認識率を損ねる場合があった。これを解決するために、上記のように適應に用いるデータ量が少ない場合には、適應に用いるパラメータを減らして過適應を防ぐという方法があるが、この場合には、適應による効果を減少させてしまい、結果的に認識率を低下させてしまっていた。

【0009】本発明は上記の問題点に鑑みてなされたものであり、認識精度を向上することができる音声認識装置及びその方法、コンピュータ可読メモリを提供することを目的とする。

【0010】

【課題を解決するための手段】上記の目的を達成するための本発明による音声認識装置は以下の構成を備える。即ち、電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識装置であって、電話番号毎に発信者に関する発信者情報を管理する管理テーブルを記憶する記憶手段と、前記発信者の電話番号を取得する取得手段と、前記取得手段で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索手段と、前記検索手段で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適應する適應手段と、前記適應手段で適應された認識環境に基づいて、前記音声データを認識する認識手段とを備える。

【0011】また、好ましくは、前記記憶手段は、前記取得手段で電話番号が取得される毎に、前記管理テーブ

ルの該電話番号に対応する発信者情報を更新する。

【0012】また、好ましくは、前記発信者情報は、少なくとも音声データ、該音声データに信号処理を施したパラメータ列、統計的情報のいずれかあるいは任意の組み合わせを含む。

【0013】また、好ましくは、前記記憶手段は、前記発信者情報として、更に、該発信者情報の作成日を前記管理テーブルに記憶する。

【0014】また、好ましくは、前記記憶手段は、前記発信者情報として、更に、発信者が使用する電話器の種類を前記管理テーブルに記憶する。

【0015】また、好ましくは、前記パラメータ列は、少なくともパワースペクトル、ケプストラム、LPCケプストラム、MFCCのいずれかを含む。

【0016】上記の目的を達成するための本発明による音声認識方法は以下の構成を備える。即ち、電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識方法であって、電話番号毎に発信者に関する発信者情報を管理テーブルに記憶する記憶工程と、前記発信者の電話番号を取得する取得工程と、前記取得工程で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索工程と、前記検索工程で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適應する適應工程と、前記適應工程で適應された認識環境に基づいて、前記音声データを認識する認識工程とを備える。

【0017】上記の目的を達成するための本発明によるコンピュータ可読メモリは以下の構成を備える。即ち、電話回線に接続され、該電話回線を介して発信者から発信された音声データを認識する音声認識のプログラムコードが格納されたコンピュータ可読メモリであって、電話番号毎に発信者に関する発信者情報を管理テーブルに記憶する記憶工程のプログラムコードと、前記発信者の電話番号を取得する取得工程のプログラムコードと、前記取得工程で取得した電話番号に対応する発信者情報を前記管理テーブルより検索する検索工程のプログラムコードと、前記検索工程で検索された発信者情報及び前記発信者から発信された音声データを用いて、認識環境を適應する適應工程のプログラムコードと、前記適應工程で適應された認識環境に基づいて、前記音声データを認識する認識工程のプログラムコードとを備える。

【0018】

【発明の実施の形態】以下、図面を参照しながら本発明の好適な実施形態を詳細に説明する。

【0019】図1は本発明の実施形態に適用可能な音声認識装置のハードウェア構成を示す図である。

【0020】図1において、H1は音声認識結果あるいは音声認識結果より得られた応答を出力する出力装置である。出力は直接ユーザに対してなされる場合だけでなく、ネットワークあるいはバスを通じて他のシステムへ

の入力の場合もある。H2は数値演算・制御等の処理を行う中央処理装置であり、後述する本発明で実行される処理に従って演算を行う。

【0021】H3はディスク装置等の外部メモリ装置やRAM・ROM等の内部メモリといった記憶装置であり、本発明で実行される処理に必要な一時的データおよび認識対象を示した文法、音声モデル、履歴データ等が格納される。また、後述する履歴テーブルH3aも記憶装置H3に格納される。H4は電話回線との接続や電話回線を通じた音声データや各種データの送受信等の制御を行う電話回線制御装置である。

【0022】以上、音声認識装置の構成要素である出力装置H1～電話回線制御装置H4は、バスH5によって相互に接続される。

【0023】以上のハードウェア構成によって実行される本発明の処理について、図2を用いて説明する。

【0024】図2は本発明の実施形態で実行される処理を示すフローチャートである。

【0025】尚、以下に説明する処理は、電話回線が接続された時点から開始され、電話交換局より発信者の電話番号が出力されるものとする。

【0026】まず、ステップS1において、発信者が用いている電話番号を電話回線制御装置H4を通じて取得する。次に、ステップS2において、取得した電話番号が過去に用いられたことがあるか否かを記憶装置H3に格納された履歴テーブルH3aより検索する。

【0027】尚、履歴テーブルH3aは、図3に示すように、過去に用いられた電話番号と、その際の入力データすなわち履歴データで構成される。履歴データは、後述するステップS4の実行に応じて、音声データ、音声データに信号処理を施したパラメータ列（具体的にはパワースペクトル、ケプストラム、LPC (Linear Prediction Coding) ケプストラム、MFCC (Mel Frequency Cepstrum Coefficient) 等）、平均値や分散等の統計的情報などの形式で構成される。

【0028】また、この統計的情報は、音声データあるいはパラメータ列から抽出される。

【0029】例えば、SS法では無音声区間のパワースペクトル系列 {パワースペクトルベクトル系列  $s_i^{(1)} \mid 1 \leq i \leq N$ 。 ( $N$  はデータ数) } が用いられる。PMC法では無音声区間のパラメータ系列 {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq N$ 。 } が用いられる。CMS法では有音声区間のケプストラムあるいはLPCケプストラム系列 {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq N$ 。 } が用いられる。話者適応では有音声区間のパラメータ系列 {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq N$ 。 } が用いられる。また、上記データ列は平均値・分散・データ数などの統計値として用いることも可能である。

【0030】次に、ステップS3において、電話回線制

御装置H4を通じて音声データを入力する。入力された音声データの無音声区間のパワースペクトル系列を {パワースペクトルベクトル系列  $s_i^{(1)} \mid 1 \leq i \leq M$ 。 ( $M$  はデータ数) } とし、パラメータ (ケプストラム) 系列を {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq M$ 。 } とし、有音声区間のパラメータ (ケプストラム) 系列を {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq M$ 。 } とする。

【0031】次に、ステップS4において、ステップS2における検索結果およびステップS3の入力結果をもとに、発信者の認識環境の適応を行う。具体的には、SS法では {パワースペクトルベクトル系列  $s_i^{(1)} \mid 1 \leq i \leq M$ 。 } の平均パワースペクトルを求めて適応を行う。PMC法では {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq M$ 。 } より雑音モデルを作成して適応を行う。CMS法では {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq M$ 。 } よりケプストラム平均を求めて適応を行う。また、話者適応法では、 {パワースペクトルベクトル系列  $c_i^{(1)} \mid 1 \leq i \leq M$ 。 } を用いてモデル適応を行ったり、話者クラスタの選択によって適応を行う。

【0032】次に、ステップS5において、ステップS4における認識環境の適応に基づいて、ステップS3で入力された音声データを音声認識し、認識結果を出力装置H1に出力する。次に、ステップS50において、上記ステップS3～ステップS5の処理を終了するか否かを判定する。終了しない場合 (ステップS50でNO)、ステップS3に戻る。一方、終了する場合 (ステップS50でYES)、ステップS6に進む。

【0033】尚、ステップS50の処理は、必要に応じて複数回繰り返される。

【0034】ステップS6において、ステップS3で入力された音声データを元に、ステップS1で取得した電話番号に対応する履歴データを更新し、終了する。

【0035】以上説明したように、本実施形態によれば、電話番号毎にその発信者に関する情報を履歴データとして管理するので、音声認識における環境適応に用いるデータ量を十分に確保することが可能となる。従って、安定して高性能の環境適応が可能となり、認識精度を向上することができる。

【0036】尚、上記ステップS2において、履歴データとともに、その作成日時を検索し、これをステップS4において利用するような構成にしてもよい。このような構成にすることで、例えば、一定時間以上古い履歴データは用いない、あるいは新しい履歴データから順に一定量の履歴データのみを用いて環境適応を行うといったことが可能である。

【0037】また、上記ステップS3～ステップS5に

において、入力データ（{パワースペクトルベクトル系列 $s'_{(i)}$ }, {パワースペクトルベクトル系列 $c'_{(i)}$ }, {パワースペクトルベクトル系列 $c'_{(i)}$ }等）が得られるのに合わせて、ステップS4において使用する履歴データの重みを減少させてもよい。このための技術として、MAP (Maximun Apriori Probability) 推定を利用することが可能である。

【0038】また、上記ステップS2において、履歴データとともに、発信者の電話番号に対応する付随的データ（例えば、電話器の種類）を検索し、これをステップS4において利用するような構成にしてもよい。このような構成にすることで、付随的データに応じた環境適応が可能となる。例えば、電話器が携帯電話やPHSの場合、その背景雑音は発信者の移動に合わせて変動するため、背景雑音に関する履歴データは用いず、回線・マイクロホンに関する環境適応のみを行うといった効率の良い環境適応を行うことが可能である。

【0039】また、一つの電話器を用いる発信者が限定される場合には、話者適応の適応度を強めることが可能である。

【0040】また、上記実施形態では、ステップS3・ステップS4・ステップS5の各ステップは、順時実行されるように説明しているが、一定時間間隔で（フレーム単位で）実行するような構成にしてもよい。

【0041】また、本発明では、背景雑音・回線特性・マイクロホン等に適応する技術は、話者認識においても適用することができる。

【0042】尚、本発明は、複数の機器（例えばホストコンピュータ、インタフェース機器、リーダ、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【0043】また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読出し実行することによっても、達成されることは言うまでもない。

【0044】この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現するこ

とになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【0045】プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

【0046】また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0047】更に、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0048】

【発明の効果】以上説明したように、本発明によれば、認識精度を向上することができる音声認識装置及びその方法、コンピュータ可読メモリを提供できる。

【0049】

【図面の簡単な説明】

【図1】本発明の実施形態に適用可能な音声認識装置のハードウェア構成を示す図である。

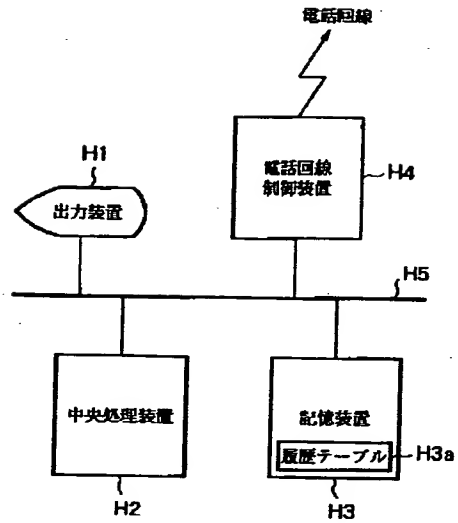
【図2】本発明の実施形態で実行される処理を示すフローチャートである。

【図3】本発明の実施形態の履歴テーブルを示す図である。

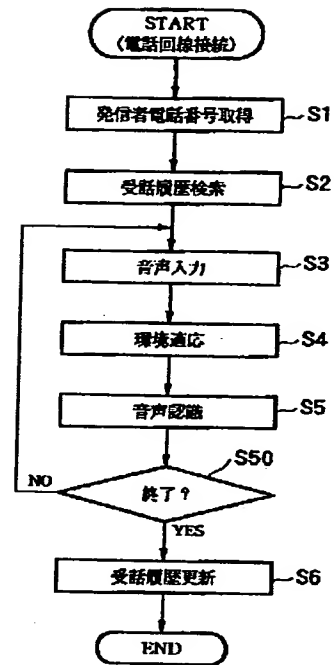
【符号の説明】

- H1 出力装置
- H2 中央処理装置
- H3 記憶装置
- H3a 履歴テーブル
- H4 電話回線制御装置
- H5 バス

【図 1】



【図 2】



【図 3】

電話番号	履歴データ
09-xxxx-xxxx	音声データ1, パラメータ列1
045-xxxx-xxxx	音声データ2, パラメータ列2
030-xxxx-xxxx	音声データ3, パラメータ列3
⋮	⋮